International Journal of Hybrid Information Technology
Vol.8, No.2 (2015), pp.389-396
http://dx.doi.org/10.14257/ijhit.2015.8.2.36

Multimode Retrieval of Breast Masses Based on Association Rules

Qian Wang[1], Yanan Lv[2] and Lixin Song[2]

[1] *School of Computer Science,*
[2]*School of Electrical and Electronic Engineering,*
Harbin University of Science and Technology
Harbin 150080, China
qiamwamg@163.com

Abstract

Breast imaging case not only image low level features but also has image semantic features. In order to implement multimode retrieval of breast imaging, using feature selection algorithm based on association rules to select features, digging out the associated relationship between image low level features and image semantic features, and then taking advantage of association classification algorithm to get image visual semantic features, which reduced the semantic gap between image low level features and visual semantic features, at last, making similarity measure combined with low level features, to make multimode retrieval come true. As the results show, this method improve the performance of breast imaging case retrieval and provide more meaningful decision support for doctors.

Keywords: *association rules, feature selection, associative classification, multi-mode retrieval*

1. Introduction

How to find the most similar case for pending inquiry case quickly and accurately from breast imaging cases database is becoming an important issue which is urgently needed to be solved. Text-based image retrieval converts image retrieval problem into traditional text retrieval problem, however, this method need a lot of manual annotation and it is highly subjective and imprecision [1].When content-based image retrieval applied in the medical field, the image low-level features extracted cannot reach the level of human understanding, there is "semantic gap"[2] between image low-level features and image semantic features, which cannot guarantee meaningful inquiry under medical background. To this end, we need a multi-mode image retrieval method combined with image low-level features and semantic features.

In recent years, in the field of image retrieval, people is paying more and more attention at the retrieval way that combined with image low-level features and image semantic features. Xie Tianhua [3] et al proposed a medical image retrieval method combined with image high-level semantic features and content low-level features, which improved the retrieval results, but it required doctors' aid descriptive semantics, and cannot get image semantic features by image low-level features. Tian Haiman *et al.* [4] used content-based hierarchical retrieval method to make computer-aided diagnosis of mass's benign and malignant by its texture, shape and boundary features come true, which get good results, but it cannot get visual semantic features of mass. Association rules can overcome this kind of deficiency, it is applied to data mining of medical imaging recent years. Wang Shuyan [5] et al used the improved Apriori algorithm for mining association rules, and established a medical image classifier, which achieved good results, Jiang Yun[6] et al constructed an enhanced

ISSN: 1738-9968 IJHIT

association rules classifier to classify medical images, which improve the accuracy of classification, however, this two methods all applied association rules to the judgment of normal and abnormal of medical cases directly, so they cannot provide visual semantic features associated with the diagnosis.

To solve these problems, we use feature selection algorithm based on association rules to select features, use mining algorithm based on association rules to obtain the association rules between image low-level features and semantic features, at the same time, reduces the dimension of low-level features. And then establish classification model by associative classification algorithm, which can obtain image visual semantic features by image low-level features, achieve machine-aided annotation, reduce the semantic gap, and on this basis, combined with low-level features to make multi-mode retrieval come true.

2. Establishment of Associative Classification Model

2.1. Related Concepts of Association Rules

Relevance of item sets can be found by association rules. Suppose I={I1,I2,I3,……}, I is item set, D is a transaction database, where each transaction T⊆I, if A is an item set, if and only if A ⊆T, we say that transaction T contains A, association is the format of A → B, where A and B all belong item set I but do not intersect. A is known as foregoing paragraph and B is known as back paragraph. Association rules have two important parameters, support and confidence. Support refers to the probability of a transaction set contains A and at the same time contains B, that is P (A ∪ B), denoted as sup, which reflects the importance of the association rules in the database; confidence refers to the ratio of the support with the probability of a transaction set contains A, that is P (B | A), denoted as conf, which measures the credibility of the association rules, namely:

$$sup = P(A \cup B) \tag{1}$$

$$conf = P(A \cup B)/P(A) \tag{2}$$

In this paper, we use classic Apriori algorithm proposed by Agrawal [7] et al. In the mining process, using class association rules mining, preceding paragraph of rules is data item sets, back paragraph is class attribute sets. Data item sets is composed by eight kinds of features, attribute sets is composed by three kinds of mass shapes, they are oval, irregular and lobulated. Each rule is represented by R, the format of rules is: R: D→C, D= {Data1, Data2⋯⋯Datan}, D is data collection, C= {C1, C2, C3}, C is class collection.

After obtaining rules by using Apriori algorithm, we need use pruning algorithm to get strong association rules. Assuming two rules R1 and R2, if one of them satisfy any conditions as following, we called R1 has the priority level than R2:

1. The confidence of R1 is higher than R2, conf (R1)>conf (R2);
2. If conf (R1) = conf (R2), but the support of R1 is higher than R2, sup (R1)>sup (R2);
3. If conf (R1) = conf (R2) and sup (R1) =sup (R2), R1 has fewer items than R2.

Pruning method in this paper is: choose rules with high priority covering low priority, if they have the same priority, choose rules with more preceding paragraph, and then get strong association rules. Using strong association rules to establish associated classification model which can do classify training to the data set.

2.2. Features Selection

In this paper, we applied StARMiner algorithm based on association rules to for mining association rules between image low-level features and shape semantics and benign and malignant, achieved the purpose of reducing dimensions and effectively associated low-level

features with semantics. Let *T* be a medical data set, *X* is a collection of image classes, X_i is an image, *fi* is the i-th feature of X_i , $\mu_{fi}(X)$ is the mean and $\sigma_{fi}(X)$ is the variance of *fi* in image *X*. This algorithm has three thresholds defined by the user, they are γ_{min} , $\Delta\mu_{min}$ and $\Delta\sigma_{max}$. γ_{min} is the lowest confidence when H_0 is not true existence; $\Delta\mu_{min}$ is the minimum difference between the mean of *fi* in *X* and other classes. $\Delta\sigma_{max}$ is the largest variance of fi in *X*. If the following three conditions are met, you can find the relationship between image *X* and features, that means feature *fi* is the key to distinguish *X*-class images and other types of images, the features should be kept. There are 32 low-level features, 8 features are kept by using StARMiner algorithm, as shown in Table 1.

$$H_0 : \mu_{fi}(X) = \mu_{fi}(T - X) \tag{3}$$

$$\left|\mu_{fi}(X) - \mu_{fi}(T - X)\right| \geq \Delta\mu_{min} \tag{4}$$

$$\left|\sigma_{fi}(X)\right| \leq \Delta\sigma_{max} \tag{5}$$

Table 1. Features

No.	Literature	Feature	No.	Literature	Feature
1	[9]	Radius curvature of segmentation region	5	[10]	Perimeter area ratio
2	[9]	Shape parameter ratio	6	[10]	Standardization radius length standard deviation
3	[9]	Standardization center offset	7	[10]	Entropy histogram normalization radius length
4	[10]	Roundness mass	8	[11]	Average length of the radius

2.3. Associative Classification Algorithm

Associative classification algorithm is developed on the basis of association rules, we take advantage of ACE [12] (Associative Classification Engine) algorithm. When building the associated classification model of image shapes, firstly, choose image low-level features of training images, use minimum length description algorithm to discrete features, and then, use Apriori algorithm for mining association rules, get strong association rules by pruning algorithm based on the degree of interest, at last, use ACE to achieve associative classification model.

ACE algorithm has four parameters: namely A (h), F (h), N (h) and ω_{min} , the formula of confidence is as follows:

$$\omega = \frac{4A(h) + F(h)}{4A(h) + F(h) + N(h)} \geq \omega_{min} \tag{6}$$

In which, w indicates the credibility of the image that belongs to a certain category, four parameters are as follows:

(1)A (h) is the numbers of image features that satisfy the entire rules;

(2)F (h) is the numbers of image features that satisfy some rules;

(3)N (h) is the numbers of image features that do not satisfy the rules;

(4) ω_{min} is the minimum value of credibility that one image belongs to a certain category.

2.4. Establishment of Breast Masses Shape Classification Model

There is correlation between benign and malignant of breast masses and different shapes of masses, for example, regular types such as oval generally appears benign, but irregular types such as irregular and lobulated appear malignant commonly. Therefore, when building classification model using image low-level features associated with shape, we make dichotomous according to the possibility of benign and malignant of different shapes, and then make detailed classifications for each node, as Figure 1.

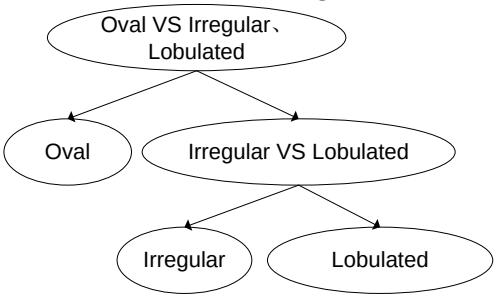


Figure 1. Shape Model

In this paper, we have 153 images, 106 images participated in image data mining, and 47 images used to test, the accuracy of the model is Table 2.

Table 2. Classification Precision

Shape	Oval	Irregular	Lobulated	Total
Associative classification precision	0.8182	0.8500	0.8750	0.8511

2.5. Retrieval System

Model retrieval system proposed in this paper is Figure 2:

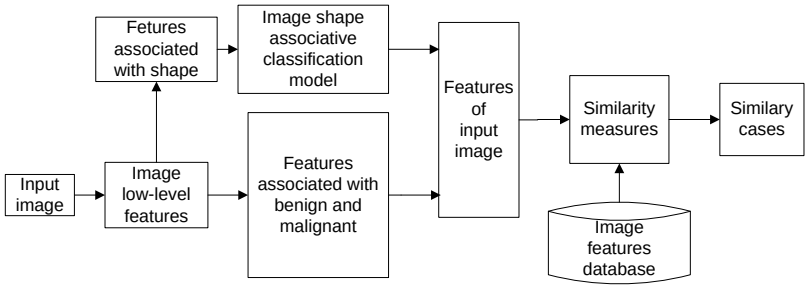


Figure 2. Retrieval System Model

At first, obtain the low-level features of the input image, choose the features associated with shape and benign and malignant, the features are Table 2. Secondly, obtain classification association rules by shape features selected, build image shape classification model by association classification algorithm, we can get shape semantic of input image by this model.

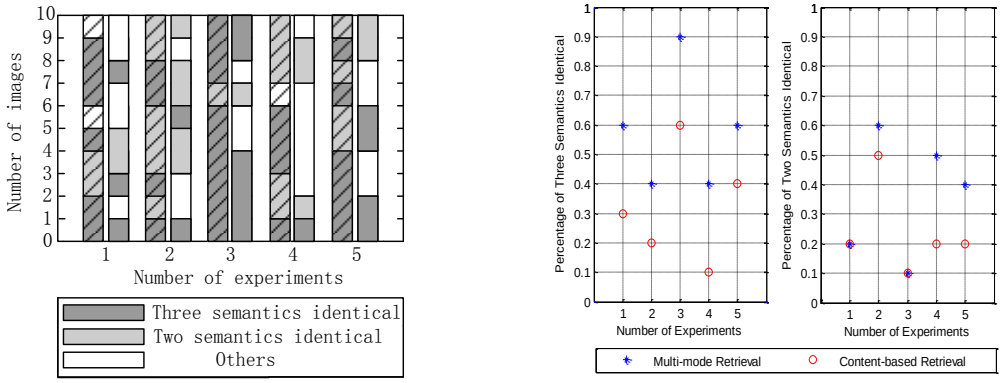
At last, combine low-level features related with benign and malignant, make similarity measure with features database, and then get the most similar image to the input image.

3. Retrieval Results

Use Euclidean distance to make image retrieval, formula as (7). The method of this paper is to determine the shape of mass, make shape semantic annotation, and then choose images that conform to the shape, conduct similarity measure, and achieve image multi-mode retrieval finally. Features used in this paper were all normalized data, m represents the m-th feature of features.

$$D(q,p)=(\sum_{m}(q_m-p_m)^2)^{\frac{1}{2}} \tag{7}$$

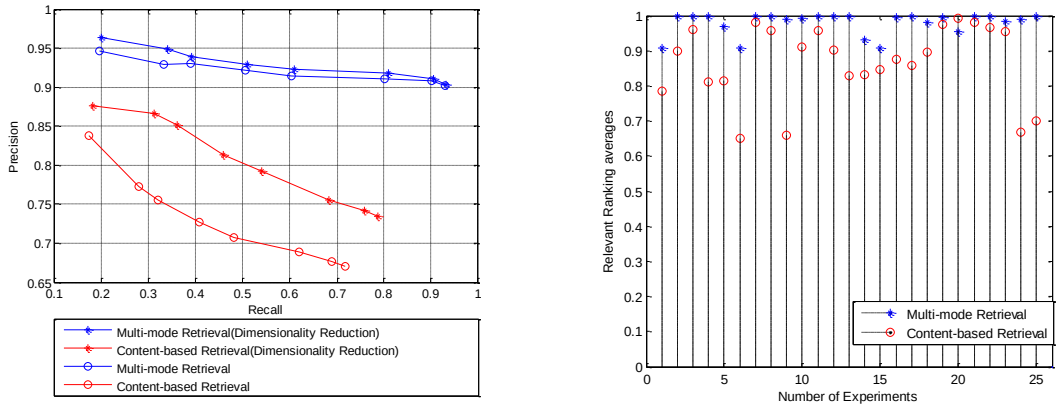
Retrieval system contains 153 images, in order to compare the effect of method proposed in this paper and content-based image retrieval in mass information retrieval, choose 5 images to make retrieval randomly, and compare the former 10 images, the shape semantics involve in the retrieval, margin and margin and benign participate in the evaluation of performance, the results shown in Figure 3. Figure 3 (a) is the comparative results sorted by semantic similarity, the results with stripes are obtained by method proposed by this paper, and the results with streak-free are obtained by content-based image retrieval method; Figure 3 (b) is the comparative results of the percentage of semantic. As can be seen from Figure3, the method proposed in this paper can provide more cases of identical semantic and cases of similar semantics for the same image.



(a) Semantic similarity (b) Percentage of semantic

Figure 3. Comparison of Semantic Similarity

The ultimate aim is to provide decision support for doctors when diagnosing the benign and malignant. We use recall-precision and relevant ranking averages to evaluate the performance of content-based image retrieval method and method proposed in this paper, comparative results shown in Figure 4. Figure 4 (a) is the curse of recall-precision, it can be seen from the figure that the precision and recall obtained by method proposed in this paper are higher than content-based retrieval, and the performance after feature selection is better than before. Figure 4 (b) is the relevant ranking averages of two methods after reducing dimensions, the larger of it, the better of effect, the relevant ranking averages of method proposed in this paper are higher than content-based retrieval. We can the method of this paper have better results from Figure 4.



(a) Curse of recall-precision (b) Relevant ranking averages

Figure 4. Comparative Results of Performance

4. Conclusion

A multi-mode retrieval method fused image semantic features and low-level features is proposed. For one image, determine the shape firstly, and then transform into feature vector, choose images in line with the shape, and combine the low-level features associated with high-level semantic to achieve multi-mode retrieval. As can be seen from the comparative results, multi-mode retrieval based on association rules has better retrieval results in the auxiliary semantic annotation and aided diagnosis, which makes up the limitations of content-based single model retrieval lacking for information. Follow-up may consider increasing the number of samples and improving association rule mining algorithm, build associative classification model for margins, in order to improve the accuracy of shape classification and provide more effective and more comprehensive semantic information for the judgment of benign and malignant.

Acknowledgements

This work is supported by Natural Science Foundation of Heilongjiang province (F200912), Science and technology innovation fund of Harbin (2010RFXXS026).

References

[1] L. Zhang, "Large-scale internet image retrieval and pattern mining", SCIENTIA SINICA Information is, vol. 43, no. 12, (**2013**), pp. 1641-1653.
[2] C. Wen and G. H. Ge, "Review and research on "semantic gap" problem in the content based image retrieval", Journal of Northwest University (Natural Science Edition), vol. 35, no. 5, (**2005**), pp. 536-540.
[3] T. H. Xie, W. J. Tang, Q. F. Zhao and J. A. Zhao, "Medical image retrieval combined image advanced semantic features with content low level features", Journal of Biomedical Engineering, vol. 26, no. 6, (**2009**), pp. 1237-1240.
[4] H. M. Tian, J. L. Lin, K. Chen and Y. L. Peng, "Content-based grading retrieval of breast tumor ultrasound images", Journal of Sichuan University (Engineering Science Edition), vol. 44 (S1), (**2012**), pp. 177-181.
[5] S. Y. Wang, M. Q. Zhou and G. H. Geng, "Research on association rule mining method for medical image", Computer Applications, vol. 25, no. 6, (**2005**), pp. 1408-1409.
[6] Y. Jiang, Z. H. Li, Y. Wang and L. B. Zhang, "A new medical image classification method based on enhanced association rules", Journal of Northwestern Polytechnical University, vol. 24, no. 3, (**2006**), pp. 401-404.
[7] R. Agrawal and Srikant, "Fast Algorithms for Mining Association Rules", Proceedings of the 20thInternational Conference on Very Large Databases (VLDB'94), (**1994**) Santiago: Morgan Kaufmann Publisher.

[8] P. H. Bugatti, M. X. Ribeiro and A. J. M. Traina, "Content-based Retrieval of Medical Images by Continuous Features Selection", 21st IEEE International Symposium on Computer-Based Medical Systems, vol. 82, (**2008**).
[9] B. Zheng, A. Lu, A. Lara, J. H. Sumkin, C. M. Hakim, M. A. Ganott and D. Gur," A method to improve visual similarity of breast masses for an interactive computer-aided diagnosis environment", Medical Physics, vol. 33, no. 1, (**2006**), pp. 111-117.
[10] H. Petrick, H. P. Chan, D. T. Wei, B. Sahiner, M. A. Helvie and D. D. Adler, "Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification", Med. Phys., vol. 23, no. 10, (**1996**), pp. 1685-1695.
[11] B. Meng, "Computer-aided Detection of Mammographic Masses Using Content-based Image Retrieval", Huazhong University of Science and Technology, (**2007**).
[12] M. X. , P. H. Bugatti, C. Traina Jr, P. Marques, N. A. Rosa and A. J. Traina, "Supporting content-based image retrieval and computer-aided diagnosis systems with association rule-based techniques", Data & Knowledge Engineering, vol. 68, no. 12, (**2009**), pp. 1370-1382.